DOCUMENT RESUME

ED 216 313 CS 006 627

AUTHOR Langer, Judith A.
TITLE The Construction of Meaning and the Assessment of Comprehension: An Analysis of Reader Performance on Standardized Test Items.
PUB DATE [82]
NOTE 38p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Secondary Education; Expectation; *Language Processing; *Reading Comprehension; *Reading Research; Reading Skills; *Reading Tests; Semantics; *Standardized Tests; *Test Interpretation; Test Items
IDENTIFIERS *Schemata; Text Factors

ABSTRACT

Using a text semantic framework, a two-year study examined the process by which readers construct meanings from standardized reading comprehension tests used as texts. The text analytic system described "ideal reader" interpretations of texts, allowing comparisons with real reader interpretations. Specific analyses described the ideal reader's hypotheses, expectations, and questions about the genre, content, and linguistic aspects of the text. A notation system was developed that provided a map of the changing envisionments that result from interpretations of earlier text segments and take new shapes as the text continues. Intensive interview procedures, including process probes, were used to compare real readers' and ideal readers' behaviors in constructing meaning for a text, as well as a reader's construction of meaning and performance on passage-related multiple choice questions. Readers of a variety of ages participated in the study, with the major analyses focusing on 26 third grade students responding to one text. The findings permitted identification of certain aspects of test language that can interfere with comprehension, including density of ideas, overreliance on assumptions and hypotheses, and assertions contrary to readers' beliefs. The findings also indicated that caution should be used when interpreting standardized test results. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED216313

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

THE CONSTRUCTION OF MEANING AND THE ASSESSMENT OF COMPREHENSION: AN ANALYSIS OF READER PERFORMANCE ON STANDARDIZED TEST ITEMS

Judith A. Langer
University of California, Berkeley

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Judith A. Langer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS806627

Background

The study discussed here used a "text semantic" approach to investigate the processes readers use to construct an understanding of test items selected from among the norm referenced standardized reading comprehension tests most frequently used by schools across the United States (note 1). The broad goals of our project were 1) to identify aspects of test language and structure that might interfere with comprehension, 2) to describe the skills and knowledge necessary for a reader to comprehend each passage, and 3) to examine the extent to which young readers have that knowledge and those skills. Every aspect of our project -- the formulation of our research questions, our system of text analysis, and our interpretation of the data -- was deeply influenced by our view of reading comprehension as a dynamically constructive process (Anderson, 1977; Bartlett, 1982; Goodman & Goodman, 1978; Polanyi, 1966; Schank & Abelson, 1977; Rumelhart, 1977).

We conducted our inquiry in two phases. The first phase focussed on the development of a systematic procedure for the text semantic analysis of test items. Our analysis codifies the lexical, syntactic, and rhetorical structures occurring in a specific text as well as the linguistic and conceptual demands these structures presumably pose for a reader (Fillmore, 1981b). It also tracks the manner in which the genre, the content, and the linguistic material of the text shape the developing meaning.

This system of text analysis served as the basis for the next phase of the study, an investigation of readers' comprehension and question-answering strategies. Using a variety of interview procedures and metacognitive probes, we gathered data which permitted us to analyze aspects of test items (each consisting of a paragraph and accompanying multiple-choice questions) that

posed difficulties for actual readers. We compared readers as meaning-integraters with readers as question-answerers in order to determine whether the test items we examined identified as high scorers the same students we identified as successful meaning-makers.

The project itself can be characterized as a series of descriptive studies carried out in the tradition of linguistic inquiry whereby our observations of readers' performance were used to inform our developing analyses of texts and readers, and these in turn were validated using other test items and other readers. All our generalizations, therefore, are an outgrowth of these repeated, in depth observations.

Ideal and Actual Readers

The analysis of a particular text proceeds linearly, tracing the changing interpretations that an "ideal reader" could justifiably make in the process of comprehending the text. The system conveys a dynamic and idealized version of reader-text interaction in that it codifies a finite number of cognitive expectations and integrations warranted by a particular text; these interpretations constitute a comprehensively developing envisionment of text meaning up to any specific point within the text.

The ideal reader, then, is an abstraction of that knowledge and those skills required for a particular interpretation of a particular text. The ideal reader is equipped with the array of schematic knowledge that the text presupposes for its interpretation, and lacks (but is prepared to learn) the material that the text introduces -- material that the reader is not yet presumed to know. The kinds of schemata or knowledge structures utilized by the ideal reader for any particular text are evoked by 1) the concepts or ideas expressed, 2) the overall structure and organization which may

rely on such devices as conversational inference, indirect speech acts, and a story grammar, and 3) the internal language (the grammar in its broadest sense, including the lexicon). Our analysis of the envisionment constructed by the ideal reader portrays the manner in which these knowledge structures operate and interact in the developing meaning of a passage.

This notion of ideal reader is not to be confused with a "mature" or "real" reader. Every reading experience, even among fully competent readers, is somewhat idiosyncratic, especially since good readers may use somewhat different processing operations to arrive at a range of acceptable interpretations. Real life reading experiences also differ from our "ideal" in that readers have varying ways of interacting with text depending on their purpose for reading. For example, an actual reader does not necessarily want to make a complete interpretation when quickly reading a "whodunnit" mystery, but most certainly does want to make a fuller interpretation when preparing for an essay examination. Our analysis does not account for such alternative purposes for reading (although theoretically it could). The ideal reader we have created for this project is a very specific idealized reader designed to exemplify the kinds of cognitive operations deemed appropriate and useful in processing specific texts. (For a more complete examination of the ideal reader concept, see Fillmore, 1981a and Kay, 1981.)

Our procedures for text analysis focus on the lexical, syntactic, and rhetorical structures that occur in consecutive meaning segments, the structures that contribute to the developing "envisionment" of the meaning of that particular passage. Codification permits us to locate the schemata and the schematic links an ideal reader draws upon or needs to construct in order to envision what is happening at any point in the text. Through this

procedure we are also able to see where these processes are interrupted or thwarted by the text. When the ideal reader is left with unanswered questions, our theory judges the text to be defective -- in some way incoherent or inadvertently misleading. In turn, we can use our analysis to compare how real readers depart from the ideal reader in developing an envisionment of the meaning of a particular passage; we can identify gaps in a reader's world knowledge or strategic knowledge -- gaps that have inhibited comprehension of the text.

The System of Analysis

As a passage is processed schemata are evoked and then integrated as the processing continues. Meaning derived from any given portion of the text is shaped by how earlier segments were interpreted and continues to develop and change in light of later segments. These changing "envisionments" are a record of the "text internal" world that is constructed by the reader while processing the text. These envisionments are the primary "dynamics" through which the reader experiences the "message." Our system of analysis keeps track of the envisionments throughout the reading experience. It also keeps track of the cognitive processes executed in the reader's effort to construct these envisionments.

Since any text includes some concepts that are explicitly stated and many others that require varying degrees of inference, we distinguish among a variety of levels of envisionment that correspond to the degree of text-based or reader-based influence. Although greater or fewer distinctions are possible in categorizing levels of envisionment, we have most recently been working with four levels:

E_0 - The most literal level of the envisionment, including only those elements that are directly warranted by what the text says;

E_1 - That level which requires text-based inferences to be made;

E_2 - That level which requires inferences triggered by the text, but which is based on personal knowledge or assumptions;

E_3 - That level of the reader's "text world" that is not warranted by the text but represents the idiosyncratic embellishments of an actual reader. (E_3 by definition does not occur in our analysis of the ideal reader.)

Our analysis of a text keeps track of the changing levels of envisionment constructed by the ideal reader. It also specifies the schemata an ideal reader might use to construct, at a particular point in time, an acceptable envisionment. We can also follow the envisionments by actual readers to compare these with ideal readers' constructions.

Domains of analysis. For our analysis we have distinguished three sources which simultaneously influence the developing envisionment: genre, content, and text.

Genre (Gn) concerns itself with readers' hypotheses about the kind of text they are reading. A genre hypothesis is generally made early in the reading experience (as early as the first sentence) and influences a reader's expectations about what the text will "say" and how it will "end."

Content (Co) refers to the evolving base of information and events.

Text (Tx) refers to the grammatical and rhetorical aspects of the text; these features create relationships that help readers raise questions and resolve hypotheses.

In responding to elements in these three domains, readers perform a number of different types of cognitive operations in the process of generating meaning. Our system of analysis tracks six general operations:

1. Questions (Q) -- uncertainties the reader has at any point during reading.

2. Hypotheses (H↓) -- predictions the reader makes about what the genre is, about what the function of a particular piece of text is, or about the answer to a question, based on a specific portion of the text.

(H→) -- predictions the reader makes about what will be "said" in succeeding portions of the text.

3. Assumptions (Ass) -- meanings the reader takes for granted without textual evidence.

4. Schemata (Sch) -- basic memory structures evoked about genre, content, or text.

5. Conclusion (Con) -- information which substantiates a hypothesis.

6. Validation (Val) -- proof that a hypothesis (H↓) was correct or a (H→) fulfilled.

Additional processes were identified in some of our notations and our system was refined as the project progressed. (For a detailed presentation of our system of analysis and notation see Fillmore and Kay, 1980 and 1982.)

Figure 1 contains an abbreviated example of the opening segment of one test passage annotated using our system (note 2). The left column presents

- - - - - - - - - - - - - - - - - -

Insert figure 1 about here

- - - - - - - - - - - - - - - - - -

our ideal reader analysis of the text segment, tracing in detail the sources

BRONCO BUSTER

(1) If a bronco buster wants to win a rodeo contest,
LO(1) If-clause, present tense, indefinite articles

Text Analysis	Related Interview Questions
1. Sch (Co) CONTEST PARTICIPANTS, DESIRE TO WIN, COMPETITION, RULES, CRITERIA FOR DECIDING ON WINNERS...	1. What do you think happens in a contest?
2. Sch (Co) RODEO CONTESTANTS, AUDIENCE, EVENTS, RIDING, ROPING, TYING ...	2. What kinds of things happen in a rodeo?
3. SCHEMATIC LINKS a RODEO presents a number of CONTESTS	
4. Sch (Tx) N+N COMPOUND N^1 identifies a type of N^2	4. What do you think they meant by "rodeo contest?"
5. Sch (Co) BRONCO BUSTING HORSE: WHEN WILD WILL TRY TO THROW RIDER RIDER: RIDES WITHOUT GETTING THROWN, BY BREAKING WILL OF HORSE RENDERS HORSE TRAINABLE	5. What is a bronco buster? What does a bronco buster do?
6. SCHEMATIC LINKS BRONCO BUSTING is one of the kinds of events in a RODEO	
7. Sch (Tx) N+V-er COMPOUND compound designates someone who V's and N's	7. ("What does somebody who is a pig-washer do? What about a mailbox painter?")
8. Sch (Tx) CONDITIONAL SENTENCE ANTECEDENT, CONSEQUENT	
9. Hyp↓ (Gn) GENRE-EXPOS PROSE	9. What kind of passage do you think this is? Do you think it might be the kind of thing you might read in a story book?..etc...
10. Hyp↓ (Th) THEME=WINNING RODEO CONTEST	10. What do you think this passage is going to be about?
11. Hyp→(Tx) the next clause will be a modal clause, expressing obligation	11. What do you think the next couple of words are going to be?
12. Hyp↓ (Tx) the two noun phrases "a bronco buster" and "a rodeo contest" are to be interpreted as generic	12. From what we know so far, is this story going to be about how to be a bronco buster or about one bronco buster in particular?
13. Ass (Co) E^1 when bronco busters choose to enter rodeos, they do so because they want to win	13. If bronco busters decide to be in a rodeo contest do you think they want to win, or doesn't it make a difference?

Figure 1

of the developing envisionments. In turn, this detail allowed us to develop a comprehensive set of questions to probe the developing envisionments of real readers confronted with these texts. The questions generated for this particular segment are listed in the right hand column in figure 1.

Procedures

Subjects

During the two years of this study, many texts were selected from an array of standardized tests and analyzed using the procedures described above. Readers of a variety of ages were then interviewed following the general format described below. Over the course of the study, a continuing pattern of text analysis followed by observation of real readers was used to refine both the text analysis system and the interview procedures. This effort also provided us with a wide range of examples of how envisionments develop (or go astray) across a text; these shaped our interpretation of the specific set of data reported below, and will be drawn upon in some detail in the general discussion of results.

The major analyses to be reported here are based on intensive study of the developing envisionments of 26 third graders in response to the test item "Bronco Buster" (figure 2). The students were selected from a middle class elementary school in Oakland, California and have a varied ethnic and home language background. Because our focus was on comprehension of test items, students with decoding problems were not included in our sample. Results of the California Test of Basic Skills, Grade 3, were available for all students. Their percentile mean for the vocabulary sub-

- - - - - - - - - - - - - - - - -

Insert figure 2 about here

- - - - - - - - - - - - - - - - -

If a bronco buster wants to win a rodeo contest, he must observe the contest rules. One of these rules is that the rider must keep one hand in the air. A rider who does not do this is disqualified.

1. A bronco buster who ignores the rules is

A. skillful B. disqualified C. chosen D. winner

2. In a rodeo contest a bronco buster must keep one hand

A. under B. still C. free D. hold

Figure 2: Bronco Buster

test was 70.5; their comprehension mean was 73.3.

Interviews

The text was presented to students in meaning segments corresponding to those used in describing the developing envisionment of the ideal reader. After a student had read a segment, questions such as those in the right hand column of figure 1 were asked. Because readers derive some meaning from clues in the prior text, and from such features as sentence length, paragraph length, and text length, each segment appeared with all previously read segments showing, and unread segments indicated but concealed. When reading the second segment of Bronco Buster, for example, the students were presented with the item illustrated in figure 3. This procedure necessitated

- - - - - - - - - - - - - - - - -

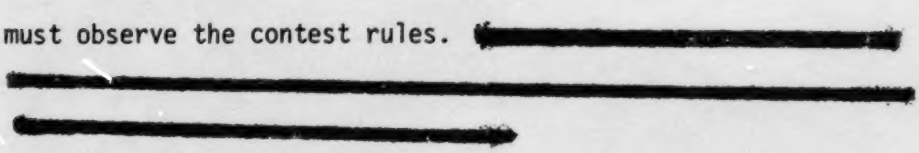
Insert figure 3 about here

- - - - - - - - - - - - - - - - -

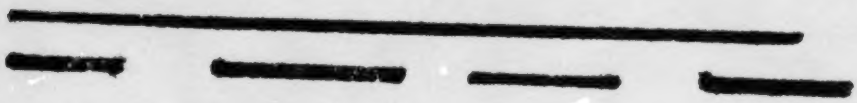
that each short text be presented in a packet of several pages before it could be completely revealed.

Since text segmentation of this type may introduce variables that differ from those a student might encounter when reading the same text in an unmutilated version, two different interview procedures were developed. Half the students began by reading an unmutilated version of the text, followed by an oral retelling. They were then presented with the mutilated version and its accompanying questions, segment-by-segment. The other half of the students began with the segmented text and accompanying questions, without the opportunity to read and retell the text as a whole. Interviews for both groups began with a training session to familiarize the students with the mutilated text format, and ended with a general question about the "main point" of the story. Results from the two procedures were essentially

If a bronco buster wants to win a rodeo contest, he must observe the contest rules.



1.



2.

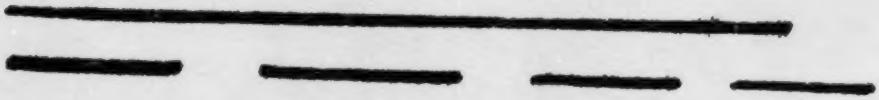


Figure 3: Bronco Buster

identical and will be discussed together in the sections below.

Each interview session lasted between 45 minutes and one hour. Each session was taped and transcribed. The transcripts were coded in terms of specific aspects of the reader's developing envisionment, and evaluated to determine each student's overall ability to construct a reasonable, evolving envisionment of the meaning of the text. These analyses permitted us to make comparisons across groups of children as well as to describe in depth the performance of individuals.

Results and Discussion

The data will be reported and discussed under two general headings: 1) observations about test takers; and 2) observations about tests. A final section will highlight some problems to consider when writing or taking tests, or using their results.

Responses to Questions

This section will contain a review of the Bronco Buster questions and responses; what answers the students expected, what answers they chose, and why (see figure 2 for the complete test item). After reading each passage segment-by-segment, students were asked to read and answer the first question, without being shown the multiple-choice selections. They were then shown the choices and asked to make a selection. Also, they were requested to explain why they "liked" the word they selected and did not "like" the words they did not select. (For an analysis of how test questions relate (or do not relate) to the passage, see Coleman, 1982.)

Responses to the two Bronco Buster questions present an interesting contrast: 20 of the 26 students eventually selected the correct answer from among the four alternatives in each item, but the answers they

expected before seeing the choices were quite different.

Question 1: A bronco buster who ignores the rules is

A. skillful B. disqualified C. chosen D. winner

For the first question, even before seeing the choices, 14 students (54%) anticipated that "disqualified" would be the correct response, eight were unable to answer, and four gave other responses. All 14 who anticipated "disqualified" also selected it as the correct response, as did another six students who had not originally anticipated that response. Of the six who selected incorrect responses to this item, four chose "winner" and two chose "skillful."

Because our interview questions are keyed to important aspects of the reader's developing envisionment, we can locate quite precisely the comprehension processes that led some readers to select the wrong responses. In this case, most of the problems stemmed from lack of knowledge of two words: "disqualified" (which appears in this passage and in the question) and "ignores" (which appears only in the question). Although the 14 children who anticipated "disqualified" were familiar with its meaning, those who selected it from among the other choices had difficulty pronouncing it, were uncertain of its meaning, and selected it either because it seemed to be a key word in the passage or simply due to the visual match from passage to response. Of the four students who chose "winner," two explained that if you do not follow the rules you'll cheat and thereby increase your likelihood of winning. Although they did not know what "disqualified" meant, a combination of context clues and world knowledge led them to construct a reasonable rationale for "winner" as the correct response. The

remaining four students were not familiar with either "ignore" or "disqualified" and chose "skillful" or "winner" based on their knowledge of a bronco buster's prowess. For at least those four readers, uncertainty about the meaning of the question was an obvious impediment to their selection of the correct response.

If the test-maker intended the question to measure students' ability to gain the meaning of a difficult word ("disqualified") from the context of the passage, the inclusion of another difficult word ("ignores") in the question thwarted this goal.

Question 2: In a rodeo contest a bronco buster must keep one hand

A. under B. still C. free D. hold

After reading the stem to the second question, 17 students (65%) anticipated that "in the air" would be the response, a plausible conclusion to the stem although it does not appear as one of the choices. Five students who anticipated "in the air" got this question wrong when forced to select from the alternatives given; four chose "hold" and one chose "still." (The other incorrect response was also "hold" -- by a student who did not offer any anticipated response.)

For this test item, the students' explanations about the plausibility of each response are especially revealing. Although finally rejecting it, two children explained that "under" was a logical response, reasoning that the bronco buster has to keep one hand in the air and the other hand under the rope. Four children considered selecting the word "still," reasoning that if you do not hold your hand still you will fall off. One child stated that if you hold one hand in the air the other one has to be still.

Eight students considered the word "hold" at some point because if you don't hold you'll fall off. Two youngsters reasoned that if you keep one hand in the air you'll have to hold with the other one. The 26 readers did not all have a complete envisionment of the passage. Some did not know what a bronco buster was (interpreting bronco as a model name for an automobile), some confused the person with the animal, and still others were uncertain as to the kind of animal (cow, bull) a bronco was. However, all youngsters demonstrated some understanding that a bronco buster keeps one hand in the air and holds on to something with the other hand (in the interview they were asked to act this out). All of their explanations of possible responses indicated an understanding of this concept, even when they got the question itself wrong.

It seems, in this case, that the youngsters who selected an incorrect response did so for reasons reflecting an accurate construction of the meaning of this part of the passage. The range of possible responses permitted at least three semantically defensible selections. Of course, the issue of grammaticality needs to be addressed. In choosing among the alternatives provided, the readers we observed in this study focussed more of their attention on selecting a word with appropriate meaning than on appropriate grammaticality within the sentence. Students sometimes admitted their selections didn't "sound right" but selected them anyway because they had the "right meaning." Fillmore (Note 2) suggests that young children and bilingual individuals may be accustomed to hearing constructions with which they are unfamiliar and therefore, even when aware of the "ungrammaticality" of a response, may be less inhibited by

that aspect than by imprecise meaning in choosing a response.

After the taped interviews were transcribed and analyzed, a coding sheet was completed for each student. At least two raters reviewed the categories to which each response was assigned and also supplied a holistic rating of each student's envisionment on a four point scale. Table 1 compares these envisionment ratings (based on detailed understanding of the passages) with performance on the questions that followed. The

- - - - - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - - - - -

results suggest that there is a fairly high association between good envisionment and response accuracy as well as between poor envisionment and poor test performance; the trend is in the right direction.

From these analyses it becomes apparent that individual readers may select incorrect test item responses based on interpretations that are contextually acceptable; these readers never have the opportunity to demonstrate their understanding of the passage. Sometimes this is because the questions miss central aspects of the envisionment; in other cases it is a function of the idiosyncratic array of response items. However, those readers who are better meaning integrators also tend to be those who select the predetermined correct responses. Findings from our analyses suggest that although the test questions themselves do not measure the integration of meaning, the ability to develop a good envisionment and the ability to select a correct response are in some way related.

Table 1

Envisionment	Number of students	Number with incorrect responses	Mean % incorrect responses
good	9	1	6
okay	11	5	23
fair	3	2	33
poor	3	2	67

Relationship between incorrect test responses and envisionment rating

Considerate and Inconsiderate Test Items

A test item is inconsiderate if it makes unwarranted demands on a reader, if it creates great cognitive demands without corresponding conceptual payoff. This is not to suggest that all difficult texts are inconsiderate and all easy ones considerate, but that the effort put forth in reading a hard text needs to be rewarded. It also suggests that the cognitive demands must be reasonable for the intended audience, the implied purpose, and the particular topic. This section will present five aspects of inconsiderate tests, aspects that have been apparent in both our "ideal reader" text analyses and our analyses of real readers reading real test items. In general, inconsiderate test items create gaps between the reader and the text that are not directly related to how well the reader constructs meaning, but are inherent instead in the way the test-maker composed the passage or presented the question. Although the five factors are presented as discrete categories, in actual tests any number of them may be interwoven; one inconsiderate feature often leads to the emergence of another.

1. Density of Ideas

Test items can be inconsiderate in the number of ideas that are presented and the frequency with which these ideas change. The brief passages used in test items sometimes provide too little elaboration of one concept before another concept is introduced. The ideas are too short-lived and change too quickly for the reader to develop any adequate envisionment. Strange Machine (figure 4) is an example of such a text (Note 4). The ideal reader analysis of this item portrays a rapid succession of new schemata

- - - - - - - - - - - - - - - - -

Insert figure 4 about here

- - - - - - - - - - - - - - - - -

In 1877 a machine appeared which surprised many people. Can you guess the name of this strange new machine?

As you spoke into the mouthpiece and turned the handle, a tube covered with a thin piece of tin moved arund. As the tube moved, a needle pressed deep lines into the tin. As you turned the handle once more, the needle touched against the same lines and played back your words.

This was the first phonograph! How different from the Hi-Fi of today!

- -

9 About how many years ago did the machine first appear?

1) 25 2) 50 3) 100 4) 200

10 How did people feel when they saw this new machine?

5) angry 6) astonished

7) worried 8) frightened

11 As the handle was turned, which part of the machine moved?

1) tube 2) needle

3) press 4) mouthpiece

12 As the needle pressed lines into the tin, it was

5) giving a shot

6) recording a voice

7) painting a picture

8) sewing a piece of material

Figure 4: The Strange Machine

introduced without any integration into a larger schematic whole. This density is deliberate in this passage, inherent in the riddle format and postponed answer about the identity of the strange machine. The reader is asked to envision a mouthpiece, which is associated with a range of possible schemata. Before the schema relevant to this text can be instantiated, another concept, "handle," is introduced. The ideal reader questions the array of possible "meanings" Q(Co) and forms a rapid succession of hypotheses (H↓Co) about the relationships between the parts of the machine. Rather than finding resolutions to these hypotheses and questions, the density of the text continues to build and the "riddle" remains unsolved until the answer is provided at the end of the passage.

While the ideal reader notation portrays a series of unresolved questions and hypotheses, real readers make judgments more rapidly even when little assistance is provided by the text. In our interviews, some children (third graders) developed an initial envisionment of the "mouthpiece" and clung to it, forcing each bit of new information into their initial envisionment. One reader envisioned a coke machine with a "mouthpiece" for the coins and a "handle" as the drink selector. The "tube" (introduced next in the passage) became the round slot from which the tin coke can was ejected. Others tried to change their envisionments to accommodate the new concepts, but created a series of disconnected images instead. One reader envisioned a mailbox, then turned it into a garage mail slot (with the slot as mouthpiece), and when this no longer worked, envisioned a lawnmower with the grass catcher as mouthpiece, the switch as the handle, and the blades as the tube and needle. Still others moved from concept to concept without even attempting to integrate the content.

The density of the ideas introduced without elaboration is counter-productive to the development of a cohesive envisionment. Hypotheses about the nature of the strange machine (which would integrate the various separate schema) are thwarted by another aspect of the text. The ideal reader assumes, when the riddle is posed at the end of the second sentence, that the "strange machine" will be similar enough to its modern version to be guessable (Ass.; Genre). In fact, the early version of the machine is not at all like the present version, as the last line of the passage emphasizes. The assumption of similarity contributes to a series of resolutions (Res) which invalidate earlier hypotheses, e.g., that the strange machine is a tape recorder, since the machine recorded and played back your words. This particular test item was inconsiderate due to the density of ideas, and the lack of resemblance between the early prototype and present phonograph rendered the passage even more difficult for the reader.

2. Overreliance on Assumptions and Hypotheses

During our investigation we encountered several test items which required the ideal reader to make a large number of assumptions (Ass), (Co), (Tx) and hypotheses (H→), (H↓), (Co), (Tx) without corroborating textual evidence to permit their validation. The envisionment is constructed from implications drawn from the ideal reader's instantiation of unarticulated referents and knowledge of social conventions. While this can be a powerful stylistic device that works well for texts such as mysteries and gothic romances, the device requires space as well as a hospitable setting. When many assumptions and unresolved hypotheses are presented in a text, uncertainty builds as new suppositions are linked with old ones, and the reader is induced to conjure up a strained envisionment of vague images.

The text in figure 5 (Note 5), for example, may at first glance appear

- - - - - - - - - - - - - - - - -

Insert figure 5 about here

- - - - - - - - - - - - - - - - -

simple for a fifteen year old to integrate. Yet, to build an envisionment of the meaning of this passage, the reader must rely on many assumptions and unresolved hypotheses. The ideal reader must make hypotheses about the antecedents for many referents throughout the passage, beginning in the very first sentence with "a voice," in the second sentence with "the reader" and "her listener," and so on throughout the text. An assumption must be made, based on social conventions, that the teenage girls were reading a "dirty" book, of which they feared the mother would disapprove. An assumption must also be made that the girl and her mother are talking across each other, each concerned with different issues: the girl with her forbidden book, and the mother with her forbidden floor-sitting. The reader must also infer that the girl is about fifteen. The actions (reading the book, hiding it, rising, glancing at the bed) are all linked by inference. This text type is inconsiderate due to the many premature hypotheses the reader must make; usual and appropriate inference strategies become tentative and strained.

3. Imitation Genre

There are a variety of genre that consistently appeared in the standardized reading tests we reviewed: stories, personal accounts, folk tales, letters, poems, and expository passages containing information about social studies, science, or health. Their inclusion seems based on the view that a reading comprehension test should include items from the variety of text types that students normally encounter in school. However, our analysis

"Alice!" called a voice.

The effect on the reader and her listener, both of whom were sitting on the floor, was instantaneous. Each started and sat rigidly intent for a moment; then, as the sound of approaching footsteps was heard, one girl hastily slipped a little volume under the coverlet of the bed, while the other sprang to her feet and in a hurried, flustered way pretended to be getting something out of a tall wardrobe.

Before the one who hid the book had time to rise, a woman of fifty entered the room and, after a glance, cried, "Alice! How often have I told you not to sit on the floor?"

"Very often, Mommy," said Alice, rising meekly, meantime casting a quick glance at the bed to see how far its smoothness had been disturbed.

"And still you continue such unbecoming behavior."

"Oh, Mommy, but it is so nice," cried the girl. "Didn't you like to sit on the floor when you were fifteen?"

- -

1. Alice's companion was

   A. A girl
   B. Her brother
   C. The family dog
   D. A doll

Figure 5

of the items disclosed that the passages sometimes violated the conceptual or structural patterns typically associated with the genres they were meant to reflect. In our analyses both of the ideal reader and of real readers, this led to genre-based hypotheses (Hyp; Gn) that were later invalidated. Although sometimes the effects on the envisionment were minor, in other cases the deviations from expected genre patterns caused major interference. For example, a short humorous story generally ends with its punchline. If any written material follows the punchline, it is likely to elaborate the "joke;" it can be read quickly, for embellishment. The reader familiar with this genre knows this convention and does not expect important information to follow. However, some test items place key information after the punchline, or ask questions dealing with what may take place afterwards; because this violates a convention of the genre, the reader with knowledge of that convention is placed at a disadvantage. Figure 6 is an example of this test and question type (Note 6). Instead of completing the action

- - - - - - - - - - - - - - - - - -

Insert figure 6 about here

- - - - - - - - - - - - - - - - - -

with the punchline, this test items requires the reader to predict what Fritz will do next. This caused some readers to doubt their interpretation of the punchline and therefore to select response A instead of B.

The passage in figure 7 (Note 7) is another example of a test item constructed to imitate a real genre, in this case expository prose. Unlike the genre the passage is imitating, however, there is no introduction

- - - - - - - - - - - - - - - - - -

Insert figure 7 about here

- - - - - - - - - - - - - - - - - -

to help the reader make a link from personal knowledge to new information,

Fritz lived in a neighborhood with many interesting people. But Fritz felt out of place. He felt that he was ordinary. In fact, he felt so ordinary that he thought nobody noticed him. Today I'll be different, he thought. I'll wear an extra hat.

Then he went for a walk. People smiled at him. But no one spoke. At last he asked a neighbor, "Don't you see something unusual about me today?"

"Yes," the neighbor said, "You're wearing three hats instead of your usual two."

- -

What will Fritz probably do next?

A. Ask more people if they think his hat looks different?

B. Realize that people have always noticed him

C. Go to the store and buy some more hats

D. Stop going for walks in his neighborhood

Figure 6

The nose cone of a rocket carries the payload. The payload is anything men wish to send into space.

The payload can be a satellite. A satellite is a small object that travels around a larger object in space. The rocket travels fast to carry the satellite away from the earth.

The payload can be a camera. It can take pictures of the earth or the moon. The payload may be a man, shooting through space.

- -

16 Which of the following can a space camera be used for?

5) to take picture of friends
6) to measure the temperature
7) to take pictures of the moon
8) to measure the speed of rockets

17 The nose cone carried the

1) fuel
2) payload
3) rocket
4) airplane

18 A small object that travels around a larger one is called a

5) camera
6) satellite
7) payload
8) nose cone

Figure 7

and the passage ends abruptly without a statement tying the passage together--all conventions of informational genre. Here, the departures from normal genre conventions do not mislead readers, but leave them without needed context for constructing an envisionment. The reader must begin with some quickly accessible knowledge about nose cones and rockets in order to even attempt to develop a cohesive envisionment. There is not a sharing of information about rockets from writer to reader, but a simple "listing" of key points. The reader who is familiar with the topic will be able to recognize these points, elaborate and connect them, and make sense of the passage, while the reader who is not already familiar with the topic will be at a loss. (Here we see again how one kind of problem, based on the use of an abbreviated genre form creates another, the kind of density of ideas discussed earlier.)

We can see how this can interfere with a developing envisionment in the responses of a third grader interviewed as he read the passage. In the first sentence, he knew what rockets looked like (from television) and thought he knew what a nose cone was. He was not certain, but expected that it would be elaborated later in the text. He had no schematic links between rockets and payload, although he knew from the sentence structure that they must have some connection. In the third sentence, he recognized the word satellite and knew it had something to do with outer space, but he had no further knowledge of it. The fourth sentence did not help him because he could not envision what an object in outer space would be like.

Although this student was trying to make meaning from the passage and had a good idea of the information he would need to know in order to comprehend, in this case the text was uncooperative; it did not contain

the kinds of structure and elaboration generally provided in informational texts. It could make sense only to a reader who already knew enough about rockets, satellites, and payloads to make the relevant connections among them.

A test item is not a literary work nor an actual piece of informational writing. Test writers seem to have other goals in mind than trade book authors; their purposes and constraints are different. Test items must comply with a particular reading level, be topically appropriate for specific age groups, contain information which can be answered by specific questions, be inoffensive to all possible groups, and so on. In an attempt to mimic the variety of genre encountered in school and also to meet these "testing" constraints, test passages might be thought of as "imitation" genre, almost genre unto themselves with conventions of their own. It is this "imitation" which renders them inconsiderate: they invoke conventions which they do not follow.

4. Assertions Contrary to Readers' Beliefs

Yet another aspect of inconsiderate text occurs when test items make assertions contrary to readers' beliefs without acknowledging differences or providing schematic links from reader knowledge to textual content. Instead, some texts we analyzed required the ideal reader to make immediate hypotheses (H↓, H→) based on an acceptance of the discordant assertion.

Figure 8 (Note 8) is an example of such a test item, from a fifth grade test. In the ideal reader notation of the text segment "while attendance at baseball games has been falling off," there is a content hypothesis Hyp (Co) that falling attendance implies loss in popularity. There is also a

- - - - - - - - - - - - - - - - -

Insert figure 8 about here

- - - - - - - - - - - - - - - - -

Games of great speed and constant motion, such as football, basketball, and ice hockey, have been gaining in popularity, while attendance at baseball games has been falling off. This is largely because the game of baseball is slower and its action irregular. That is, the game alternates between great moments of high drama and excitement -- such as when the bases are loaded and the batter hits a home run -- and extended periods of comparative inactivity, such as when a batter takes full count, two strikes, and three balls, and then is walked on the last pitch. Many ideas are being explored for making baseball a faster game. One possibility, for example, might be to increase the distance between the pitcher and the batter, thereby allowing a split second more for the batter to judge the pitch and connect with the ball. This could produce more hits, thereby making baseball more of a running rather than a pitching game.

Figure 8

text hypothesis Hyp↓(Tx) that "while" contrasts baseball's falling attendance with football, basketball, and ice hockey which are gaining in popularity. There are neither schematic nor text links elaborating this idea. The next text segment ("This is largely because") goes on to explore explanations of the loss in popularity. In this segment the ideal reader makes two hypotheses: Hyp↓(Tx) that "this" refers to baseball's falling attendance, and Hyp↓(Tx) that "largely" suggests more than one reason for the falling attendance. This is followed by a hypothesis Hyp↓(Tx) in the next text segment that baseball is "slower" than other games.

Real readers interviewed with this text found the assertion about baseball's loss in popularity jarring. Since children see baseball games and baseball players in a wide variety of media contexts, they assume it is a very popular game. During an interview, one fifth grader reader exhibited well-developed passage envisionment. However, when choosing the response to a question about the decrease in baseball's attendance, he said angrily, "I know people are going to come to baseball games. I've gone to other people's houses and they watch. And I've seen the stands just stuffed. They can barely cram another person in." Another reader knew the individual word meanings of "popularity" (like a movie star, very common), "falling off" (not coming to), and "attendance" (to come to). Despite this, his envisionment was poorly integrated. It wavered between baseball's gaining and not gaining in popularity, and therefore baseball's relation to the other games also wavered. Both readers appeared unable, on their own, to consistently link the text assertion with their personal

beliefs about baseball attendance. This form of inconsiderate text results when the author fails to acknowledge the conflicting ideas that might arise, and the text fails to encourage the reader to question personally held ideas or beliefs that are contrary to those assumed in the text.

5. Contextual Discontinuity by Deceptive Simplicity

The fifth type of inconsiderate text occurs when a concept with which a reader is unfamiliar is presented bit-by-bit in a sequence of seemingly familiar details. In the ideal reader analysis, integration of the details leads to an envisionment of the whole concept. The notation of the Strange Machine test item (see Figure 4) is an example). In this passage, the strange machine is described bit-by-bit and the parts can be integrated as follows (partial notation):

As you spoke into the machine and turned the handle,

e^1 - the clause describes the use of the machine in 1877
Sch(Co) mouthpiece: part of machine to do with the voice
Sch(Co) handle: part of machine - helps it to function when turned
Schematic link - machine with mouthpiece and handle, something to do with voice communication

a tube covered with a thin piece of tin moved around.

Q (Co) what is relationship of tube to mouthpiece and handle?
Q (Tx) how is the tube moving around?
H_o(Co) the tube is also part of the machine
e^o - in the strange machine, as you speak into the mouthpiece and turn the handle, the tube turns

As the tube moved, a needle pressed deep lines into the tin.

H(Co) the deep lines are related to the voice function
Schematic links - the machine has a needle, mouthpiece, handle, and a voice-related function
H(Co) the lines are pressed to record a message

As you turned the handle once more

H(Co) the machine does more than press lines
e^{o} - to make the machine work properly the handle must be turned at least twice

the needle touched against the same lines and played back your words.

Sch(Co) voice recording
Val - a voice recording is made
Val - the lines have a role in the machine's functioning
Con E^{o} - this strange machine allowed people in 1877 to record their voices
H(Co) this was a tape recorder? record player?

This was the first phonograph!

Con - the parts described are similar enough to the modern day phonograph to call it the first phonograph

How different from the hi fi of today!

Con - both the phonograph and hi fi play back sounds and are thereby related to the strange machine

As we have already seen in discussing density of ideas, this passage leaves the reader with a high level of unintegrated detail. To further complicate the matter, the vocabulary is kept deceptively simple, deceptively because the concepts introduced have a wide range of possible referents (e.g., crank, bar, switch, lip for handle) and can therefore be attributed to a large number of schemata. Because the conceptual whole is not provided by the text, the reader must develop an envisionment by relating the bits to attributes which are part of an existing textually plausible schema. However, an integration problem occurs if the reader does not already have (or is unable to access) an appropriate schema, if the text concept is so remote that it cannot reasonably be related to known schemata, or if the text doesn't provide clues to help the reader select among the possible schematic choices.

This problem was evident in our interviews when it became apparent that although the students were familiar with tape recorders, the handle, tube, and needle did not fit into their notion of a prototype tape recorder. And although they were familiar with record players, their notion of a prototype record player did not include a personal voice-recording function. Further, the third graders we interviewed were not familiar with either phonographs or hi fi's although they were quite knowledgeable about stereos and record players. In the absence of an apparently useful schematic structure, many students clung to their first image, as accommodating necessary to new information. For example, one student envisioned (e^3) a robot with a mouth (mouthpiece), hands (handles), and a tube with wheels to move. The tube was covered with tin "so you won't get a shock." It also pressed lines "so you can see what he says. He draws letters." In this instance these apparently simple text segments evoked schemata which were totally reader-based. This reader did not have the benefit of either text or knowledge structures to validate his hypotheses. Therefore, lack of familiarity with the larger whole to which the "simple" parts relate precludes knowledgeable selection of appropriate schemata from among the array of those which, in an abstract sense, are potentially possible.

Discussion

This study has examined some of the consequences of test language and its relationship to comprehension. It has also analyzed test passages in terms of the demands test items make on raders. The two preceeding sections identify some of the complexities which may severely confound the results obtained from standardized tests, complexities which pose processing or

question-answering problems unrelated to students' ability to construct meaning.

These data suggest the need for caution when considering the purposes for the administration of tests and what it is that is being tested. If the goal is to make large scale discriminations between better and poorer school achievers, reliance on linguistic and conceptual "puzzles" may suffice. However, if the purpose is to measure reading comprehension, to discriminate among the range of more and less successful users of meaning construction strategies, a test specifically designed to examine these strategies must be devised. Our data indicate that although the reading comprehension tests we examined do discriminate good from poor reading, the strategies required for success on these items do not bear a strong relation to the processes involved in constructive meaning-making. Standardized reading tests do not measure the processes involved in the construction of meaning from a text nor do they evaluate an individual's ability to manage those processes.

References

Anderson, R.C. The notion of schemata and the educational enterprise. In R.C. Anderson, R.J. Spiro & W.E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, New Jersey: Lawrence Erlbaum, 1977.

Bartlett, F.C. Remembering. Cambridge, England: Cambridge University Press, 1932.

Coleman, L. Answering test questions: Idealized readers in new test situations. Working paper, Reading Project, Language Behavior Research Laboratory, University of California, Berkeley, 1982.

Fillmore, C.J. Ideal readers and real readers. Georgetown University Round Table, 1981.

Fillmore, C.J. On monitoring and describing the reading process. Paper presented at National Reading Conference, Dallas, Texas, 1981.

Goodman, K.S. & Goodman, Y.M. Reading of American children whose language is a stable rural dialect of English or is a language other than English. Final report to the National Institute of Education, 1981.

Kay, P. Three properties of the ideal reader. Georgetown University Round Table, 1981

Polanyi, M. The tacit dimension. Garden City, New York: Doubleday, 1966.

Rumelhart, D.E. Notes on a schema for stories. In D.G. Bobrow & A.M. Collins (Eds.), Representation and understanding: Studies in cognitive science. New York: Academic Press, 1975.

Schank, R.C. & Abelson, R.P. Scripts, plans, goals and understanding. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1977.

Reference Notes

1. The work reported here was supported, in part, by the National Institute of Education under grant number G-790121 Rev. 1, "Text Semantic Analysis of Reading Comprehension Tests." Co-principal investigators were Charles J. Fillmore and Paul Kay; research team members were Karen Carroll, Linda Coleman, Katherine Kovacic, and Mary Catherine O'Connor. Welcome advice was received from Mary Sue Ammon, Kjel Madsen, Haj Ross, and Patrizia Violi.

2. Fillmore, C.J. Discussion during project meetings, 1981.

3. Gates-MacGinitie Reading Tests, Primary C Form 1. New York: Teachers College Press, 1965.

4. Comprehensive Test of Basic Skills, Level 1 Form S, CTB. Monterey, California: McGraw-Hill, 1974.

5. Cooperative English Tests. Princeton, New Jersey: Educational Testing Service, 1960.

6. SRA Achievement Series, Level E Form 1. Chicago, Illinois: Science Research Associates, 1978.

7. Comprehensive Test of Basic Skills, Level 1 Form S, CTB. Monterey, California: McGraw-Hill, 1974.

8. Stanford Achievement Test, Intermediate II Form A. New York: Harcourt, Brace and Jovanovich, 1972.